REMARKS

Claims 1-23 remain in the application. An IDS is filed herewith to illustrate knowledge and concepts that were well-known in the prior art.

Claim Rejections - 35 U.S.C. §112

As an initial matter, Applicants understand that the term "efficient" as used in the claims might raise concerns in regard to M.P.E.P. 2173.05(b) as a possible use of relative terminology. First, Applicants note that the term is not a relative term such as those, like "about," "essentially," "similar," "substantially," "type," "relatively shallow," or "of the order of" discussed in the M.P.E.P. It is a descriptive term.

Secondly, as explained in further detail below, the term "efficient" has an understood meaning in the art and the specification provides standards, in the form of examples, that would apprise one of skill in the art as to the claim scope. The Examiner is further reminded that breadth of a claim is not to be equated with indefiniteness (see M.P.E.P. 2173.04).

Claims 4 and 16 - "efficient structure"

Claims 4 and 16 recite the claim limitation of "packing said data into an efficient structure." The Office Action has objected to the term "efficient structure" in this limitation as being allegedly vague and indefinite. To support this position, the Office Action cites to the *Merriam-Webster* dictionary definition of "efficient" as "productive of desired effects" to conclude that "one of skill in the art would question what characteristics must be present for the structure to contain these desired effects."

As previously presented by Applicants, the present specification lists examples of an "efficient structure" in the field of bioinformatics. The term does not need an explicit definition since its meaning is readily understood by one of skill in the art. As a term of art in bioinformatics, the Office Action's citation to *Merriam-Webster* is inappropriate. Indeed, in the background portion of the present specification, Applicants discuss the well known use of NCBI BLAST's *formatdb* program to pack sequence data into one of the natural and commonly used efficient structures (2-bits per nucleotide) disclosed at paragraph 19:

> "The NCBI BLAST suite of programs includes a program called *formatdb* that converts a text-based sequence database into a specialized format. The *blastall* program requires such a formatted database for the subject database when it performs a search. The "*blastn*" algorithm implemented in *blastall*, for instance, requires a formatted nucleotide database for the subject database. In order to reduce the size of the database, and as an optimization for speedier comparison of nucleotide sequence data, *formatdb* converts each ASCII character representing a single nucleotide (e.g. 'A', 'C', 'G', 'T', 'U', 'a', 'c', etc.) into a two bit value representing one of the four standard nucleotides, A, C, G or T(U). These are then packed four at a time into 8-bit bytes in a packed database file."

Likewise, the cited art of record to Matsumoto et al. discussing "Biological Sequence Compression Algorthms" states at the bottom of page 43 that "[s]ince DNA sequences contain four symbols, 'a,' 't,' 'g,' and 'c,' if these were totally random, *the most efficient way to represent them would be using two bits for each symbol.*" (Emphasis added)

As noted by the Matsumoto et al. paper, compression schemes exist that take advantage of the non-random nature of DNA to provide further compression. However, by providing examples in the specification to the use of the minimum number of bits needed to express the possibilities, Applicants submit they have provided one of skill in the art with the minimum characteristics required of an "efficient structure," thus circumscribing the claimed metes and bounds, albeit in a broad manner.

And finally, Altschul et al. *do*, in fact suggest that those of skill in the art understand efficient packing since Altschul et al. disclose that it "is advantageous to compress the database by packing 4 nucleotides into a single byte" (p. 405, col. 1), which, because a byte consists of 8 bits, is the same as 2-bits per nucleotide.

In view of the subjective documentary evidence, Applicants submit that the term "efficient structure" is definite to one of skill in the art and that claims 4 and 16 comply with the second paragraph of 35 USC 112.

Claims 7 and 19 - "efficiently encoded representation of alignment"

As with claims 4 and 16 above, the Office Action alleges that a claim term is vague and indefinite based upon the term "efficiently," with a further statement that

Applicants "have not shown where in the specification [using the minimum number of bits needed to represent the data] is explained or sufficiently proven (via documentation) that this definition of the phrase is well known in the art."

In response, the Applicants submit that the well known *formatdb* program which was disclosed as follows in the specification:

> "to reduce the size of the database, and as an optimization for speedier comparison of nucleotide sequence data, *formatdb* converts each ASCII character representing a single nucleotide (e.g. 'A', 'C', 'G', 'T', 'U', 'a', 'c', etc.) into a two bit value representing one of the four standard nucleotides, A, C, G or T(U),"

as discussed in paragraph 19 uses efficient encoding. Furthermore, the disclosure of Matsumoto et al. states that:

> "[s]ince DNA sequences contain four symbols, 'a,' 't,' 'g,' and 'c,' if these were totally random, the *most efficient way to represent them* would be using two bits for each symbol"

and thus sufficiently demonstrates that one of skill in the art understands what is meant by "efficiently encoded" sequence data.

Furthermore, additional evidence that the concept of efficient encoding is well understood in the art of bioinformatics is hereby submitted. Varre et al. is a 1999 article published in *Bioinformatics* and, with respect to encoding scripts for sequence comparison, discloses:

> "To be comparable, descriptions have to be written in the same language. We use binary language *because efficient encoding procedures are known.* As DNA is made up from 4 ($=2^2$) possible bases, each of them might be encoded over 2 (the exponent) bits. A n-bases long sequence is thus encoded over $2n$ bits." (Emphasis added)

In view of the subjective documentary evidence, Applicants submit that the term "efficiently encoded representation of alignment" is definite to one of skill in the art and that claims 7 and 19 comply with the second paragraph of 35 USC 112.

Claim 8 - "seed point and sum set membership"

The Office Action further alleges that "seed point and sum set membership" is vague and indefinite since "it is unclear how the Applicants intend this phrase to be

defined." In rejecting Applicants arguments, the Office Action alleged that Applicants' prior arguments were not supported by documentation and the terms were not used by Altschul et al.'s original BLAST paper.

In response, Applicants agree with the implicit tenets of the Office Action that one of ordinary skill in the art would be aware of the work of Altschul et al. and would be familiar with NCBI's BLAST program. Indeed, as shown in the attached Karlin et al. reference, Altschul and Karlin discussed the concept of "sum" statistics in 1993 to address multiple high-scoring segments pairs (HSPs) in molecular sequences that can occur due to gaps in "consistently ordered segment pairs in sequence alignments." The present invention discloses the use of BLAST 2 (2.0.14), which has sum statistics output enabled as a default, as shown in the "Search Strategy" portion of the attached BLAST Help Manual, wherein it says:

> "By default the programs use 'Sum' statistics (Karlin and Altschul, 1993). As such, the statistical significance ascribed to a set of HSPs may be higher than that ascribed to any individual member of the set. Only when the ascribed significance satisfies the user-selectable threshold (E parameter) will the match be reported to the user."

Since the sum statistics for an alignment relate to a set of HSPs comprised of various members, the identified HSPs were referred to by the present inventors and others in the field as the "sum set" and membership of an HSP therein as "sum set membership." The terms that appear to be the most often used in the art are "*set of HSPs*," as in the above quote from the BLAST Help Manual, and "*sum group*," as used in note 6 of the Release History section of the BLAST 2.0 Release Notes, attached hereto. However, the present application's use of "sum set" is well understood in the art to be synonymous with "sum group" and "set of HSPs" and therefore is sufficiently definite to one of skill in the art.

Likewise, "seed point" is well understood by those of skill in the art. As stated in Karlin et al. in the second paragraph under the heading "The Construction And Statistical Evaluation Of Gapped Local Alignments," a seed is a single aligned pair - "Starting from a single aligned pair of residues, called the *seed*, the dynamic programming proceeds..."

If the sequences being compared are known, the most efficient way to identify the seed is to identify the point at which it occurs in the sequence, that is the *seed point*.

Indeed, the term "seed point and sum set membership" refers to the minimum information required to reconstruct (on the client side) multiple gapped alignments, and group the ones together which give the smallest resulting summed e-value (in accordance with Karlin et al.). That is, it identifies the location of each alignment (the seed point) and the HSPs that go together as "summed" alignments (the sum set membership). In view of this, Applicants submit that the term is sufficiently definite to one of skill in the art.

Further, in response to the Office Action's statement that the term was not used in the cited BLAST prior art, such as the cited 1990 reference to Altschul et al., Applicants point out that the statistical approach of Sum statistics was not part of the original BLAST and was first introduced by Karlin and Altschul in 1993, as noted above. However, the cited prior art to Anderson et al. further demonstrates the use of the term "seed" as a term of art to refer to the sequence S(0) chosen for comparison (see, e.g., pages 350-352). Additional evidence that "seed" is a term of art used in this manner is submitted herewith in the form of the article by Brudno et al., which on page 2 states that the sequence comparison algorithm "works by chaining together pairs of similar regions, one from each of the two input DNA sequences; we call such pairs of regions *seeds*. More precisely, a seed is a pair of words of length k with at least n identical base pairs."

In view of the subjective documentary evidence, Applicants submit that the term "seed point and sum set membership" is definite to one of skill in the art and that claim 8 complies with the second paragraph of 35 USC 112.

Claims 1 and 13 - "said task definition"

Claims 1 and 13 were rejected in the Office Action based upon an alleged lack of antecedent basis for "said task definition." However, Applicants submit that use of the singular form "said task definition," when the plural terms "tasks" and "task definitions" have been claimed, clearly limits the antecedent basis to the prior singular limitation "sending a task definition for each task from the master CPU to one of a plurality of slave

CPUs" in claim 1 and "sending a task definition for each task to one of said plurality of slave CPUs" in claim 13. In view of this, "said task definition" clearly refers to *each* task definition that is sent to *each one* of the plurality slave CPUs.

Applicants therefore submit that the term "said task definition" has proper antecedent basis in the claims and complies with the second paragraph of 35 USC 112.

In view of the forgoing arguments, Applicants respectfully request reconsideration and withdrawal of the rejections under 35 USC 112.

Claim Rejections - 35 USC 103

Claims 1, 4, 6-13 and 18-23

Claims 1, 4, 6-13 and 18-23 have been rejected under 35 USC 103 as being obvious over Smith et al. in view of Altschul et al. and Reed et al. To establish a *prima facie* case of obviousness, three basic criteria must be met (See M.P.E.P. Section 2143). First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988); *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992).

Second, there must be a reasonable expectation of success. This requirement is primarily concerned with less predictable arts, such as the chemical arts.

Finally, the prior art must teach or suggest each and every limitation of the claimed invention, as the invention must be considered as a whole. *In re Hirao*, 535 F.2d 67, 190 U.S.P.Q. 15 (C.C.P.A. 1976).

The teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, not in Applicants' disclosure. *In re Vaeck*, 947 F.2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991).

No Motivation to Combine

In the present case, none of these criteria have been met in the Office Action. First, there is no suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the search launcher interface of Smith et al. or combine it with Altschul et al. and Reed et al.

M.P.E.P. 2141.02 requires that an invention be considered as a whole. The present invention, *as a whole*, is drawn to a method or system for comparing a query dataset N to a subject dataset M using not only a network, but a *distributed computing platform*. A client computer in the claimed system and method divides the query dataset N into n_N data elements having a size within a specified range, divides the subject dataset M into n_M data elements having a size within said specified range, and determines a number of tasks for an entire comparison of datasets N and M as $n_N \times n_M$. The client computer then sends all data elements and task definitions to a master CPU of a master-slave distributed computing platform, and the master CPU sends a task definition and its associated data elements for each task to one of a plurality of slave CPUs of the distributed computing platform. The slave CPUs of the distributed computing platform perform the tasks (inherently in parallel) and return the results to the master CPU.

In contrast, *none* of Smith et al., Altschul et al. or Reed et al. even mentions distributed computing. In making the rejection, the Office Action erroneously looks to *Merriam-Webster* for the definition of "system" instead of looking to the broadest reasonable interpretation *consistent with the specification* as required by M.P.E.P. 2111.

M.P.E.P. 2141.02 further requires that the prior art be considered as a whole, including portions that teach away from the invention. Smith et al., as a whole, *teaches against* the present invention in teaching the use of a batch system that processes various sequence searches serially "one at a time" at a single site (the BCM Search Launcher server, see Abstract, lines 14-17 and page 461, column 2, discussing batch processing) instead of in parallel at multiple slave CPUs, as found in the present invention. Smith et al. is merely a client-server system for providing a search launcher WWW interface and merely provides access to existing WWW services on remote servers. No matter how the Office Action twists or mischaracterizes Smith et al. (i.e., "Smith et al. describes ... promoting a distributed information space by filling out an HTML form..."), it is a fact that neither the client nor the BCM server include any step or software for splitting up a NxM dataset comparison into $n_N \times n_M$ tasks. Likewise, it is a fact that client search requests in Smith et al. are processed serially and that each search request is sent to a single remote site. A fair reading of Smith et al. illustrates that the disclosed system is

merely a WWW gateway to pre-existing search services and that it can perform some pre-processing in the form of batch entry and post-processing in the form of adding links to results. It does nothing to solve the problems existing in the prior art, such as (1) that sequence-to-database comparisons (as illustrated in fig. 1 of Smith et al.) require large RAM requirements for efficient processing or (2) that typical BLAST queries over a network involve sending inefficient ASCII (256-bit) characters (as illustrated by the "cut and paste" sequence entry disclosed by Smith et al.).

Likewise, Altschul et al., as a whole, *teaches away* from the present invention by teaching dataset-to-dataset comparison on a single machine (i.e., "a shared memory version of BLAST...loads the compressed DNA file into memory once" is the only disclosed technical performance enhancement). Although Altschul et al. discloses the comparison of two random sequences n and m, it nowhere suggests dividing the problem further, let alone dividing it into tasks for different computers to solve, as erroneously asserted by the Office Action.

The Office Action's citation to Reed et al. borders on the ridiculous. Reed et al. has nothing to do with bioinformatics. It has nothing to do with dataset comparisons. Indeed, it has nothing to do with solving large computational problems with distributed computing (but rather information distribution). Reed et al. is drawn to an "automated communications system [that] operates to transfer data, metadata and methods from a provider computer to a consumer computer through a communications network." The disclosed compression in col. 57 is for word processing documents with PKZIP, not the databases of col. 14 as the Office Action implies. Like the other references, it *teaches away* from the present invention since, as the cited paper to Matsumoto et al. teaches on page 44, "if one applies the standard text compression software such as `compress` or `gzip`, they cannot compress DNA sequences, but only expand the file with more than two bits per symbol." The present invention applies standard redundancy reduction data compression to an efficiently packed data element to avoid this issue, not the raw ASCII data that Smith et al. and Reed et al. seek to transmit over networks.

The stated motivation for the combination in the Office Action, i.e., that "it would have been obvious one having ordinary skill in the art at the time the invention was made

to compress data (as stated by Altschul et al. and Reed et al.) and looping processes [sic] (as stated by Reed et al.) in order to offer enhanced, integrated, easy-to-use, and time-saving techniques to a large number of useful molecular biology database search and analysis services for organizing and improving access to these tools for Genome researchers worldwide (Smith et al., page 459, col 1, third paragraph to col. 2, first paragraph)" is not only incomprehensible, but it further is *completely unrelated to limitations of the claimed invention.* It is clearly an improper hindsight reconstruction, not even of the claimed invention, but merely for the purpose of combining the disparate references that the Examiner found that use appropriate words like "BLAST," "server," "network," "distributed," "database," and "compression," which apparently turned up in the required electronic text searches.

Indeed, the Office Action has completely failed at making a *prima facie* case of obviousness under *Graham v. Deere* since it has failed to identify or evaluate *any* of the differences between the claimed invention and the prior art.

No Reasonable Expectation of Success

One of ordinary skill in the art could not reasonably be expected to find Applicant's claimed invention for comparing large datasets obvious in view of a plurality of references that provide no guidance on handling large datasets or processing them in parallel over a network. Indeed, if the compression teaching suggested by the Office Action were implemented (PKZIP compression of ASCII DNA data), the network would be saturated (and fail) due to the *expanded* file sizes that would result therefrom.

All Claim Limitations Not Shown

Smith et al. teaches the running of sequence-to-database searches, but fails to teach or fairly suggest numerous claim limitations required by all of the claims, including at least:

- dividing said query dataset N into n_N data elements having a size within a specified range;
- dividing said subject dataset M into n_M data elements having a size within said specified range;
- determining a number of tasks for an entire comparison of datasets N and M as

$n_{N \times} n_M$;

- sending all data elements and task definitions to a master central processing unit (CPU) of a master-slave distributed computing platform,

  wherein task definitions comprise at least one comparison parameter, at least one executable element capable of performing comparisons, a query data element identification(ID)/descriptor, and a subject data element ID/descriptor, and

  wherein data elements are sent alternately from query and subject data elements;

- sending a task definition for each task from the master CPU to one of a plurality of slave CPUs when all parts of a task definition and data elements referenced by said task definition are available at said master CPU;

- sending data elements referenced by said task definition to said slave CPU; and

- performing each task on a slave CPU.

Selection of a sequence to "clip and paste" into the HTML input form of Smith et al. is not a *division* of a query dataset N, but rather a specification of dataset N. No datasets in Smith et al. are ever divided, no tasks (*plural* for a single N-M comparison) are determined, and no subject dataset elements are ever sent to a Master CPU.

Altschul et al. fail to disclose any of the limitations missing from Smith et al. It merely discloses the basic BLAST algorithm for sequence comparison, i.e., comparing one sequence with another sequence, or for searching a database. Like Smith et al., Altschul et al. at least fail to disclose or suggest dividing sequence comparison problems into discrete segments for processing on a plurality of CPUs, let alone any specific method of doing this task.

Reed et al. also fail to remedy any of the defects of the Smith et al. and Altschul et al. references and is completely unrelated to the present invention.

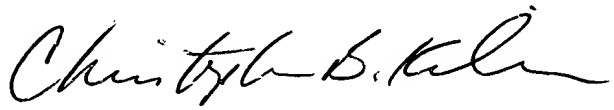
As a whole, none of the cited prior art teaches or fairly suggests dividing the problem of comparing datasets M and N into $n_{N \times} n_M$ comparisons of data elements from N with data elements from M as presently claimed. For at least these reasons, Applicant

submits that the claims are allowable over the prior art and requests reconsideration and allowance of the claims.

Conclusion

For the reasons stated above, Applicants submit that the application and claims 1-23 are in condition for allowance and respectfully requests withdrawal of all of the rejections. If there remain any issues that may be disposed of via a telephonic interview, the Examiner is kindly invited to contact the undersigned at the local exchange given below.

Respectfully,



Christopher B. Kilner
Registration No. 45,381
Roberts Abokhair & Mardula, LLC
11800 Sunrise Valley Drive, Suite 1000
Reston, VA 20191-5302
(703) 391-2900